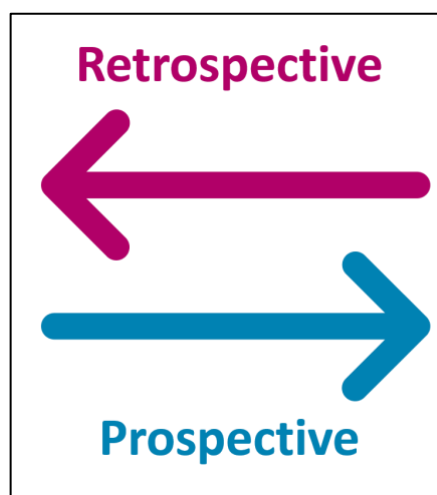


## 2 Retrospective vs prospective harmonisation

There are two main types of harmonisation:

- Retrospective harmonisation
  - We do this when we make data that have already been collected more comparable (i.e. data that already exists).
- Prospective harmonisation
  - We do this when we collect new data in a way that is intended to ensure it is comparable to data from other sources.



### 2.1 Retrospective harmonisation

Longitudinal studies can vary in important ways, such as their scientific focus, their sample, and the time periods they cover.

For examples, the MRC National Survey of Health and Development started collecting data from their cohort members in the 1940s, while the Millennium Cohort Study started their

data collection almost 60 years later. Important changes can occur over such time intervals.

Measurement practices can be revised and modified. New technologies can also emerge. Sometimes the changes occur in how the measurements are undertaken, e.g. a shift from pen and paper to computer-based questionnaires. Additionally, the way the data are recorded or stored might change due to new developments in data management software or changes to administrative processes.

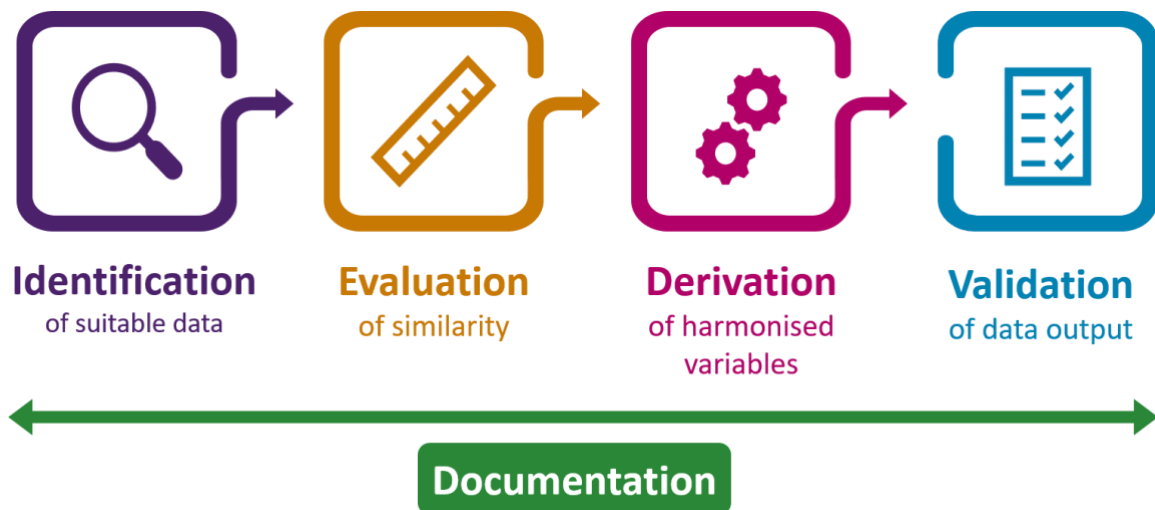
Scientific understanding can also change and develop, meaning that studies might change over time in what characteristics of their participants they measure (e.g. new questions might be added to reflect changing lifestyles or household structures). These can all impact upon the ease with which we can bring together and compare data from different studies or time periods. One approach to addressing and overcoming these challenges is through data harmonisation.

### ***How is retrospective harmonisation undertaken?***

Retrospective harmonisation involves a process of:

- Looking at what data are available,
- Evaluating their comparability and how appropriate and easy it may be to undertake harmonisation,
- Modifying the data to make them more similar and derive harmonised versions of the variables (e.g. recoding categorical variables (such as ethnicity or occupational classifications) so that the data are grouped in the same way for each sweep/study or transforming continuous variables (such as height or weight) so that they are on same scale),
- Validating the output.

Documentation underpins all of this, from accessing sufficient detail about how the variables were defined and collected originally by studies, to capturing the decisions made subsequently during the harmonisation process itself (e.g. how the data were processed and changed).



Sometimes we find that there are insufficient data available to undertake harmonisation that is both valid and worthwhile. For example, too few studies might have assessed the characteristics in which we are interested, or they might have assessed those characteristics in ways that lack any compatibility.

Even where retrospective harmonisation seems possible, we may end up having to modify the data so much to derive the harmonised variable that we lose important detail and precision and the variable may no longer be informative. That is why prospective harmonisation can offer some advantages, as explained in the next section.

## 2.2 Prospective harmonisation

If a study plans from the outset to collect data in a way that is comparable to how it was collected in an earlier sweep, or by a different study, this is called 'prospective harmonisation'. In order to sufficiently prepare for future data harmonisation, documentation from previous studies can be important to understanding details related to the assessment methods that were used.

Prospective harmonisation is not without its own challenges. It can restrict studies from using more novel assessment methods or approaches customised to their specific research question or characteristics of their sample. However, prospective harmonisation does not have to mean asking exactly the same questions of participants or running all of

Authors: Dara O'Neill, Neil Kaye and Rebecca Hardy

the same tests but only those for which comparison or data pooling is planned. Moreover, they may not have to be exactly identical. Sometimes it is possible to use slightly different assessments if there is a statistical basis for calibrating the different measures. For example, researchers have produced algorithms for adjusting measures captured by older types of blood pressure monitor to make them more comparable to blood pressure values produced by newer devices).

As documentation is at the heart of all forms of harmonisation, metadata concepts and standards play a key role. To learn more about what metadata means and why standardised documentation methods are helpful to cross-study and cross-sweep research, see **Understanding Metadata**.